

DISEASE-INDUCED POLYNUCLEOTIDES

benefit of

The present invention claims priority in part from US Provisional Application Serial No. 60/125,814 filed March 23, 1999.

5

FIELD OF THE INVENTION

This invention is in the field of plant molecular biology and relates to compositions and methods for modifying a plant's traits, in particular plant disease tolerance or resistance.

10

BACKGROUND OF THE INVENTION

Gene expression levels are controlled in part at the level of transcription, and transcription is affected by transcription factors. Transcription factors regulate gene expression throughout the life cycle of an organism and so are responsible for differential levels of gene expression at various developmental stages, in different tissue and cell types, and in response to different stimuli. Transcription factors may interact with other proteins or with specific sites on a target gene sequence to activate, suppress or otherwise regulate transcription. In addition, the transcription of the transcription factors themselves may be regulated.

15

Because transcription factors are key controlling elements for biological pathways, altering the expression levels of one or more transcription factors may change entire biological pathways in an organism. For example, manipulation of the levels of selected transcription factors may result in increased expression of economically useful proteins or metabolic chemicals in plants or to improve other agriculturally relevant characteristics. Conversely, blocked or reduced expression of a transcription factor may reduce biosynthesis of unwanted compounds or remove an undesirable trait. Therefore, manipulating transcription factor levels in a plant offers tremendous potential in agricultural biotechnology for modifying a plant's traits.

20

25

The present invention provides transcription factors for use in modifying a plant's disease tolerance or resistance.

30

SUMMARY OF THE INVENTION

In one aspect, the present invention relates to a transgenic plant comprising a recombinant polynucleotide. The recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide comprising at least 6 consecutive amino acids of a sequence selected from the group consisting of protein SEQ ID Nos. 2N, where N=1-56. And the presence of the recombinant polynucleotide alters the disease tolerance or resistance

35

of the transgenic plant when compared with the same trait of another plant lacking the recombinant polynucleotide.

In one embodiment, the nucleotide sequence encodes a polypeptide comprising a conserved domain which may be 1) a localization domain, 2) an activation domain, 3) a repression domain, 4) an oligomerization domain or 5) a DNA binding domain. In a further embodiment, the nucleotide sequence further comprises a promoter operably linked to the nucleotide sequence. The promoter may be a constitutive or inducible or tissue-active.

In a second aspect, the present invention relates to a method for altering a plant's disease tolerance or resistance. The method comprises (a) transforming a plant with a recombinant polynucleotide comprising a nucleotide sequence encoding a polypeptide comprising at least 6 consecutive amino acids of a sequence selected from the group consisting of protein SEQ ID Nos. 2N, where N=1-56; (b) selecting transformed plants; and (c) identifying a transformed plant with roots having an altered trait.

In one embodiment, the nucleotide sequence encodes a polypeptide comprising a conserved domain which may be 1) a localization domain, 2) an activation domain, 3) a repression domain, 4) an oligomerization domain or 5) a DNA binding domain. In a further embodiment, the nucleotide sequence further comprises a promoter operably linked to the nucleotide sequence. The promoter may be a constitutive or inducible or tissue-active.

In a third aspect, the present invention relates to a method for altering the expression levels of at least one gene in a plant. The method comprises (a) transforming the plant with a recombinant polynucleotide comprising a nucleotide sequence encoding a polypeptide comprising at least 6 consecutive amino acids of a sequence selected from the group consisting of protein SEQ ID Nos. 2N, where N=1-56; and (b) selecting said transformed plant.

In one embodiment, the nucleotide sequence encodes a polypeptide comprising a conserved domain which may be 1) a localization domain, 2) an activation domain, 3) a repression domain, 4) an oligomerization domain or 5) a DNA binding domain. In a further embodiment, the nucleotide sequence further comprises a promoter operably linked to the nucleotide sequence. The promoter may be a constitutive or inducible or tissue-active.

In a fourth aspect, the present invention relates to another method for altering the disease tolerance of a plant. The method comprises (a) transforming the plant with a recombinant polynucleotide comprising a nucleotide sequence comprising at least 18 consecutive nucleotides of a sequence selected from the group consisting of SEQ ID Nos. 2N-1, where N= 1-56, and SEQ ID Nos. 113-121; and (b) selecting said transformed plant.

In yet another aspect, the present invention is yet another method for altering a plant's trait. The method comprises (a) providing a database sequence; (b) comparing the database sequence with a polypeptide selected from SEQ ID Nos. 2N, where N= 1-56; (c) selecting a database sequence that meets selected sequence criteria; and (d) transforming

said database sequence in the plant. Alternatively, the database sequence can be compared with a polynucleotide selected from SEQ ID Nos. 2N-1, where N= 1-56 or SEQ ID Nos. 113-121.

In a further aspect, the present invention is method for altering a plant's trait, and the method entails (a) providing a test polynucleotide; (b) hybridizing the test polynucleotide with a polynucleotide selected from SEQ ID Nos. 2N-1, where N= 1-56 or SEQ ID Nos. 113-121 at low stringency; and (c) transforming the hybridizing test polynucleotide in a plant to alter a trait of the plant.

BRIEF DESCRIPTION OF THE FIGURES

Figures 1a-1e provide a table of exemplary polynucleotide and polypeptide sequences of the invention. The table includes from left to right for each sequence: the SEQ ID No., the internal code reference number, the transcription factor family of the sequence, particular DNA or protein fragments for each sequence, whether the sequence is a polynucleotide or polypeptide sequence, identification of the coding sequence for each full length and identification of any conserved domains for the polypeptide sequences.

DETAILED DESCRIPTION OF THE INVENTION

DEFINITIONS

A "recombinant polynucleotide" is a nucleotide sequence comprising a gene coding sequence or a fragment thereof (comprising at least 18 consecutive nucleotides, preferably at least 30 consecutive nucleotides, and more preferably at least 50 consecutive nucleotides). Additionally, the polynucleotide may comprise a promoter, an intron, an enhancer region, a polyadenylation site, a translation initiation site, 5' or 3' untranslated regions, a reporter gene, a selectable marker or the like. The polynucleotide may comprise single stranded or double stranded DNA or RNA. The polynucleotide may comprise modified bases or a modified backbone. The polynucleotide may be genomic, a transcript (such as an mRNA) or a processed nucleotide sequence (such as a cDNA). The polynucleotide may comprise a sequence in either sense or antisense orientations.

A "recombinant polynucleotide" is a polynucleotide that is not in its native state, e.g., the polynucleotide is comprised of a nucleotide sequence not found in nature or the polynucleotide is separated from nucleotide sequences with which it typically is in proximity or is next to nucleotide sequences with which it typically is not in proximity.

An "recombinant polypeptide" is a polypeptide derived from the translation of a recombinant polynucleotide or is more enriched in a cell than the polypeptide in its natural

state in a wild type cell, e.g. more than 5% enriched, more than 10% enriched or more than 20% enriched and is not the result of a natural response of a wild type plant or is separated from other components with which it is typically associated with in a cell.

A “transgenic plant” may refer to a plant that contains genetic material not normally found in a wild type plant of the same species, or in a naturally occurring variety or in a cultivar, and which has been introduced into the plant by human manipulation. A transgenic plant is a plant that may contain an expression vector or cassette. The expression cassette comprises a gene coding sequence and allows for the expression of the gene coding sequence. The expression cassette may be introduced into a plant by transformation or by breeding after transformation of a parent plant.

A transgenic plant refers to a whole plant as well as to a plant part, such as seed, fruit, leaf, or root, plant tissue, plant cells or any other plant material, and progeny thereof.

The phrase “altered expression” in reference to polynucleotide or polypeptide expression refers to an expression pattern in the transgenic plant that is different from the expression pattern in the wild type plant or a reference; for example, by expression in a cell type other than a cell type in which the sequence is expressed in the wild type plant, or by expression at a time other than at the time the sequence is expressed in the wild type plant, or by a response to different inducible agents, such as hormones or environmental signals, or at different expression levels (either higher or lower) compared with those found in a wild type plant. The term also refers to lowering the levels of expression to below the detection level or completely abolishing expression. The resulting expression pattern may be transient or stable, constitutive or inducible.

A “transcription factor” (TF) refers to a polynucleotide or polypeptide that controls the expression of a gene or genes either directly by binding to one or more nucleotide sequences associated with a gene coding sequence or indirectly by affecting the level or activity of other polypeptides that do bind directly or indirectly to one or more nucleotide sequences associated with a gene coding sequence. A TF, in this definition, includes any polypeptide that can activate or repress transcription of a single gene or a number of genes. This polypeptide group includes, but is not limited to, DNA binding proteins, protein kinases, protein phosphatases, GTP-binding proteins and receptors.

The transcription factor sequence may comprise a whole coding sequence or a fragment or domain of a coding sequence. A “fragment or domain”, as referred to polypeptides, may be a portion of a polypeptide which performs at least one biological function of the intact polypeptide in substantially the same manner or to a similar extent as does the intact polypeptide. A fragment may comprise, for example, a DNA binding domain that binds to a specific DNA promoter region, an activation domain or a domain for protein-protein interactions. Fragments may vary in size from as few as 6 amino acids to the length of the

intact polypeptide, but are preferably at least 30 amino acids in length and more preferably at least 60 amino acids in length. In reference to a nucleotide sequence "a fragment" refers to any sequence of at least consecutive 18 nucleotides, preferably at least 30 nucleotides, more preferably at least 50, of any of the sequences provided herein. Exemplary polynucleotides or polypeptides comprise a sequence provided in the Sequence Listing as SEQ ID No.1 (G1043), SEQ ID No.2 (G1043 protein), SEQ ID No.3 (G759), SEQ ID No.4 (G759 protein), SEQ ID No.5 (G185), SEQ ID No.6 (G185 protein), SEQ ID No.7 (G629), SEQ ID No.8 (G629 protein), SEQ ID No.9 (G435), SEQ ID No.10 (G435 protein), SEQ ID No.11 (G4), SEQ ID No.12 (G4 protein), SEQ ID No.13 (G1035), SEQ ID No.14 (G1035 protein), SEQ ID No.15 (G179), SEQ ID No.16 (G179 protein), SEQ ID No.17 (G28), SEQ ID No.18 (G28 protein), SEQ ID No.19 (G1241), SEQ ID No.20 (G1241 protein), SEQ ID No.21 (G19), SEQ ID No.22 (G19 protein), SEQ ID No.23 (G503), SEQ ID No.24 (G503 protein), SEQ ID No.25 (G263), SEQ ID No.26 (G263 protein), SEQ ID No.27 (G921), SEQ ID No.28 (G921 protein), SEQ ID No.29 (G1275), SEQ ID No.30 (G1275 protein), SEQ ID No.31 (G242), SEQ ID No.32 (G242 protein), SEQ ID No.33 (G1006), SEQ ID No.34 (G1006 protein), SEQ ID No.35 (G1049), SEQ ID No.36 (G1049 protein), SEQ ID No.37 (G502), SEQ ID No.38 (G502 protein), SEQ ID No.39 (G239), SEQ ID No.40 (G239 protein), SEQ ID No.41 (G555), SEQ ID No.42 (G555 protein), SEQ ID No.43 (G352), SEQ ID No.44 (G352 protein), SEQ ID No.45 (G1352), SEQ ID No.46 (G1352 protein), SEQ ID No.47 (G1089), SEQ ID No.48 (G1089 protein), SEQ ID No.49 (G553), SEQ ID No.50 (G553 protein), SEQ ID No.51 (G1221), SEQ ID No.52 (G1221 protein), SEQ ID No.53 (G580), SEQ ID No.54 (G580 protein), SEQ ID No.55 (G270), SEQ ID No.56 (G270 protein), SEQ ID No.57 (G201), SEQ ID No.58 (G201 protein), SEQ ID No.59 (G1417), SEQ ID No.60 (G1417 protein), SEQ ID No.61 (G233), SEQ ID No.62 (G233 protein), SEQ ID No.63 (G920), SEQ ID No.64 (G920 protein), SEQ ID No.65 (G867), SEQ ID No.66 (G867 protein), SEQ ID No.67 (G659), SEQ ID No.68 (G659 protein), SEQ ID No.69 (G620), SEQ ID No.70 (G620 protein), SEQ ID No.71 (G596), SEQ ID No.72 (G596 protein), SEQ ID No.73 (G511), SEQ ID No.74 (G511 protein), SEQ ID No.75 (G471), SEQ ID No.76 (G471 protein), SEQ ID No.77 (G385), SEQ ID No.78 (G385 protein), SEQ ID No.79 (G261), SEQ ID No.80 (G261 protein), SEQ ID No.81 (G25), SEQ ID No.82 (G25 protein), SEQ ID No.83 (G610), SEQ ID No.84 (G610 protein), SEQ ID No.85 (G229), SEQ ID No.86 (G229 protein), SEQ ID No.87 (G221), SEQ ID No.88 (G221 protein), SEQ ID No.89 (G186), SEQ ID No.90 (G186 protein), SEQ ID No.91 (G562), SEQ ID No.92 (G562 protein), SEQ ID No.93 (G255), SEQ ID No.94 (G255 protein), SEQ ID No.95 (G3), SEQ ID No.96 (G3 protein), SEQ ID No.97 (G713), SEQ ID No.98 (G713 protein), SEQ ID No.99 (G515), SEQ ID No.100 (G515 protein), SEQ ID No.101 (G390), SEQ ID No.102 (G390 protein), SEQ ID No.103 (G1034), SEQ ID No.104 (G1034 protein), SEQ ID No.105 (G1149), SEQ ID No.106 (G1149 protein), SEQ ID No.107 (G1334), SEQ ID No.108 (G1334 protein), SEQ ID No.109 (G1650), SEQ ID

No.110 (G1650 protein), SEQ ID No.111 (G241), SEQ ID No.112 (G241 protein), SEQ ID No.113 (G348), SEQ ID No.114 (G171), SEQ ID No.115 (G521), SEQ ID No.116 (G1274), SEQ ID No.117 (G182), SEQ ID No.118 (G1290), SEQ ID No.119 (G374), SEQ ID No.120 (G682) and SEQ ID No.121 (G501).

5 A "conserved domain" refers to a polynucleotide or polypeptide fragment that is more conserved at a sequence level than other fragments when the polynucleotide or polypeptide is compared with homologous genes or proteins from other plants. The conserved domain may be 1) a localization domain, 2) an activation domain, 3) a repression domain, 4) an oligomerization domain or 5) a DNA binding domain.

10 A nucleotide sequence is "operably linked" when it is placed into a functional relationship with another nucleotide sequence. For example, a promoter or enhancer is operably linked to a gene coding sequence if the presence of the promoter or enhancer increases the level of expression of the gene coding sequence.

15 "Trait" refers to a physiological, morphological, biochemical or physical characteristic of a plant or particular plant material or cell. This characteristic may be visible to the human eye, such as seed or plant size, or be measured by biochemical techniques, such as the protein, starch or oil content of seed or leaves or by the observation of the expression level of genes by employing Northernblots, RT PCR, microarray gene expression assays or reporter gene expression systems or be measured by agricultural observations such as stress tolerance, yield or disease resistance.

20 "Trait modification" refers to a detectable difference in a characteristic in a transgenic plant expressing a polynucleotide or polypeptide of the present invention relative to a plant not doing so, such as a wild type plant. The trait modification may entail at least a 5% increase or decrease in an observed trait (difference), at least a 10% difference, at least a 20% difference, 25 at least a 30%, at least a 50%, at least a 70%, at least a 100% or a greater difference. It is known that there may be a natural variation in the modified trait. Therefore, the trait modification observed entails a change in the normal distribution of the trait in transgenic plants compared with the distribution observed in wild type plant.

30 Trait modifications of particular interest include those to seed (embryo), fruit, root, flower, leaf, stem, shoot, seedling or the like, including: enhanced tolerance to environmental conditions including freezing, chilling, heat, drought, water saturation, radiation and ozone; enhanced resistance to microbial, fungal or viral diseases; resistance to nematodes, decreased herbicide sensitivity, enhanced tolerance of heavy metals (or enhanced ability to take up heavy metals), enhanced growth under poor photoconditions (e.g., low light and/or 35 short day length), or changes in expression levels of genes of interest. Other phenotypes that may be modified relate to the production of plant metabolites, such as variations in the production of taxol, tocopherol, tocotrienol, sterols, phytosterols, vitamins, wax monomers,

anti-oxidants, amino acids, lignins, cellulose, tannins, prenyllipids (such as chlorophylls and carotenoids), glucosinolates, and terpenoids, enhanced or compositionally altered protein or oil production (especially in seeds), or modified sugar (insoluble or soluble) and/or starch composition. Physical plant characteristics that may be modified include cell development (such as the number of trichomes), fruit and seed size and number, yields of plant parts such as stems, leaves and roots, the stability of the seeds during storage, characteristics of the seed pod (e.g., susceptibility to shattering), root hair length and quantity, internode distances, or the quality of seed coat. Plant growth characteristics that may be modified include growth rate, germination rate of seeds, vigor of plants and seedlings, leaf and flower senescence, male sterility, apomixis, flowering time, flower abscission, rate of nitrogen uptake, biomass or transpiration characteristics, as well as plant architecture characteristics such as apical dominance, branching patterns, number of organs, organ identity, organ shape or size.

Of particular interest are traits relating to increased disease resistance or tolerance of a plant, such as alterations in cell wall composition, trichome number or structure, callose induction, phytoalexin induction, alterations in the cell death response or the like. These transgenic plants may be more resistant to biotrophic or necrotrophic pathogens such as a fungus, bacterium, mollicute, virus, nematode, a parasitic higher plant or the like and associated diseases. Another desirable phenotype is a change in the overall gene expression pattern of the plant in response to disease.

1. The Sequences

We have discovered particular plant transcription factors (TFs) that are induced when plants are exposed to either biotrophic or necrotrophic pathogens.. These transgenic plants may be more resistant to biotrophic or necrotrophic pathogens such as a fungus, bacterium, mollicute, virus, nematode, a parasitic higher plant or the like and associated diseases, in particular, pathogens such as *Fusarium oxysporum*, *Erysiphe orontii* and other powdery mildews, *Sclerotinia spp.*, soil-borne oomycetes, foliar oomycetes, *Botrytis spp.*, *Rhizoctonia spp.*, *Verticillium dahliae/albo-atrum*, *Alternaria spp.*, rusts, *Mycosphaerella spp.*, *Fusarium solani*, or the like. The diseases include fungal diseases such as rusts, smuts, wilts, yellows, root rot, leaf drop, ergot, leaf blight of potato, brown spot of rice, leaf blight, late blight, powdery mildew, downy mildew, and the like; viral diseases such as sugarcane mosaic, cassava mosaic, sugar beet yellows, plum pox, barley yellow dwarf, tomato yellow leaf curl, tomato spotted wilt virus, and the like; bacterial diseases such as citrus canker, bacterial leaf blight, bacterial wilt, soft rot of vegetables, and the like; nematode diseases such as root knot, sugar beet cyst nematode or the like.

These transcription factors can be used to modulate a plant's response to disease. The plant transcription factors may belong to one of the following transcription factor families:

the AP2 (APETALA2) domain transcription factor family (Riechmann and Meyerowitz (1998) *J. Biol. Chem.* 379:633-646); the MYB transcription factor family (Martin and Paz-Ares, (1997) *Trends Genet.* 13:67-73); the MADS domain transcription factor family (Riechmann and Meyerowitz (1997) *J. Biol. Chem.* 378:1079-1101); the WRKY protein family (Ishiguro and Nakamura (1994) *Mol. Gen. Genet.* 244:563-571); the ankyrin-repeat protein family (Zhang et al. (1992) *Plant Cell* 4:1575-1588); the zinc finger protein (Z) family (Klug and Schwabe (1995) *FASEB J.* 9: 597-604); the homeobox (HB) protein family (Duboule (1994) *Guidebook to the Homeobox Genes*, Oxford University Press); the CAAT-element binding proteins (Forsburg and Guarente (1989) *Genes Dev.* 3:1166-1178); the squamosa promoter binding proteins (SPB) (Klein et al. (1996) *Mol. Gen. Genet.* 1996 250:7-16); the NAM protein family (Souer et al. (1996) *Cell* 85:159-170); the IAA/AUX proteins (Rouse et al. (1998) *Science* 279:1371-1373); the HLH/MYC protein family (Littlewood et al. (1994) *Prot. Profile* 1:639-709); the DNA-binding protein (DBP) family (Tucker et al. (1994) *EMBO J.* 13:2994-3002); the bZIP family of transcription factors (Foster et al. (1994) *FASEB J.* 8:192-200); the Box P-binding protein (the BPF-1) family (da Costa e Silva et al. (1993) *Plant J.* 4:125-135); the high mobility group (HMG) family (Bustin and Reeves (1996) *Prog. Nucl. Acids Res. Mol. Biol.* 54:35-100); the scarecrow (SCR) family (Di Laurenzio et al. (1996) *Cell* 86:423-433); the GF14 family (Wu et al. (1997) *Plant Physiol.* 114:1421-1431); the polycomb (PCOMB) family (Kennison (1995) *Annu. Rev. Genet.* 29:289-303); the teosinte branched (TEO) family (Luo et al. (1996) *Nature* 383:794-799); the ABI3 family (Giraudat et al. (1992) *Plant Cell* 4:1251-1261); the triple helix (TH) family (Dehesh et al. (1990) *Science* 250:1397-1399); the EIL family (Chao et al. (1997) *Cell* 89:1133-44); the AT-HOOK family (Reeves and Nissen (1990) *Journal of Biological Chemistry* 265:8573-8582); the S1FA family (Zhou et al. (1995) *Nucleic Acids Res.* 23:1165-1169); the bZIPT2 family (Lu and Ferl (1995) *Plant Physiol.* 109:723); the YABBY family (Bowman et al. (1999) *Development* 126:2387-96); the PAZ family (Bohmert et al. (1998) *EMBO J.* 17:170-80); a family of miscellaneous (MISC) transcription factors including the DPBF family (Kim et al. (1997) *Plant J.* 11:1237-1251) and the SPF1 family (Ishiguro and Nakamura (1994) *Mol. Gen. Genet.* 244:563-571); the golden (GLD) family (Hall et al. (1998) *Plant Cell* 10:925-936).

Producing transgenic plants with modified expression levels of one or more of these transcription factors compared with those levels found in a wild type or reference plant may be used to modify a plant's traits. The effect of modifying the expression levels of a particular transcription factor on the traits of a transgenic plant is described further in the Examples.

The polynucleotides and polypeptides are provided in the Sequence Listing and are tabulated in Figure 1. Figure 1 identifies a SEQ ID No., its corresponding GID number, the transcription factor family to which the sequence belongs, fragments derived from the sequences, whether the sequence is a polynucleotide or a polypeptide sequence, the full

length coding sequences and conserved domains. We have also identified domains or fragments derived from the sequences. The numbers indicating the fragment location for the DNA sequences may be from either 5' or 3' end of the DNA. For the protein sequences the fragment location is determined from the N-terminus of the protein and may include adjacent amino acid sequences, such as for example for SEQ ID No. 2 an additional 10, 20, 40, 60 or 100 amino acids in either N-terminal or C-terminal direction of the described fragments.

The identified polypeptide fragments may be linked to fragments or sequences derived from other transcription factors so as to generate additional novel sequences, such as by employing the methods described in Short, PCT publication WO9827230, entitled "Methods and Compositions for Polypeptide Engineering" or in Patten et al., PCT publication WO9923236, entitled "Method of DNA Shuffling". Alternatively, the identified fragment may be linked to a transcription activation domain. A transcription activation domain assists in initiating transcription from a DNA binding site. A common feature of some activation domains is that they are designed to form amphiphilic alpha helices with excess positive or negative charge (Giniger and Ptashne (1987) Nature 330:670-672, Gill and Ptashne (1987) Cell 51:121-126, Estruch et al (1994) Nucl. Acids Res. 22:3983-3989). Examples include the transcription activation region of VP16 or GAL4 (Moore et al. (1998) Proc. Natl. Acad. Sci. USA 95: 376-381; and Aoyama et al. (1995) Plant Cell 7:1773-1785), peptides derived from bacterial sequences (Ma and Ptashne (1987) Cell 51: 113-119) and synthetic peptides (Giniger and Ptashne, supra).

The isolated polynucleotides and polypeptides may be used to modify plant development, physiology or biochemistry such that the modified plants have a trait advantage over wild type plants. The identified polynucleotide fragments are also useful as nucleic acid probes and primers. A nucleic acid probe is useful in hybridization protocols, including protocols for microarray experiments. Primers may be annealed to a complementary target DNA strand by nucleic acid hybridization to form a hybrid between the primer and the target DNA strand, and then extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR) or other nucleic-acid amplification methods. See Sambrook et al., *Molecular Cloning. A Laboratory Manual*, Ed. 2, Cold Spring Harbor Laboratory Press, New York (1989) and Ausubel et al. (eds) *Current Protocols in Molecular Biology*, John Wiley & Sons (1998).

2. Identification of Homologous Sequences (Homologs)

Homologous sequences to those provided in the Sequence Listing derived from *Arabidopsis thaliana* or from other plants may be used to modify a plant trait. Homologous sequences may be derived from any plant including monocots and dicots and in particular

agriculturally important plant species, including but not limited to, crops such as soybean, wheat, corn, potato, cotton, rice, oilseed rape (including canola), sunflower, alfalfa, sugarcane and turf; or fruits and vegetables, such as banana, blackberry, blueberry, strawberry, and raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, spinach, squash, sweet corn, tobacco, tomato, watermelon, rosaceous fruits (such as apple, peach, pear, cherry and plum) and vegetable brassicas (such as broccoli, cabbage, cauliflower, brussel sprouts and kohlrabi). Other crops, fruits and vegetables whose phenotype may be changed include barley, currant, avocado, citrus fruits such as oranges, lemons, grapefruit and tangerines, artichoke, cherries, nuts such as the walnut and peanut, endive, leek, roots, such as arrowroot, beet, cassava, turnip, radish, yam, sweet potato and beans. The homologs may also be derived from woody species, such pine, poplar and eucalyptus.

Substitutions, deletions and insertions introduced into the sequences provided in the Sequence Listing are also envisioned by the invention. Such sequence modifications can be engineered into a sequence by site-directed mutagenesis (Wu (ed.) *Meth. Enzymol.* (1993) vol. 217, Academic Press). Amino acid substitutions are typically of single residues; insertions usually will be on the order of about from 1 to 10 amino acid residues; and deletions will range about from 1 to 30 residues. In preferred embodiments, deletions or insertions are made in adjacent pairs, e.g., a deletion of two residues or insertion of two residues. Substitutions, deletions, insertions or any combination thereof may be combined to arrive at a sequence. The mutations that are made in the polynucleotide encoding the transcription factor should not place the sequence out of reading frame and should not create complementary regions that could produce secondary mRNA structure.

Substitutions are those in which at least one residue in the amino acid sequence has been removed and a different residue inserted in its place. Such substitutions may be conservative with little effect on the function of the gene, for example by substituting alanines for serines, arginines for lysines, glutamate for aspartate and the like. The substitutions which are not conservative are expected to produce the greatest changes in protein properties will be those in which (a) a hydrophilic residue, e.g., seryl or threonyl, is substituted for (or by) a hydrophobic residue, e.g., leucyl, isoleucyl, phenylalanyl, valyl or alanyl; (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain, e.g., lysyl, arginyl, or histidyl, is substituted for (or by) an electronegative residue, e.g., glutamyl or aspartyl; or (d) a residue having a bulky side chain, e.g., phenylalanine, is substituted for (or by) one not having a side chain, e.g., glycine.

Additionally, the term "homologous sequence" may encompass a polypeptide sequence that is modified by chemical or enzymatic means. The homologous sequence may be a sequence modified by lipids, sugars, peptides, organic or inorganic compounds, by the

use of modified amino acids or the like. Protein modification techniques are illustrated in Ausubel et al. (eds) *Current Protocols in Molecular Biology*, John Wiley & Sons (1998).

Homologous sequences also may mean two sequences having a substantial percentage of sequence identity after alignment as determined by using sequence analysis programs for database searching and sequence alignment and comparison available, for example, from the Wisconsin Package Version 10.0, such as BLAST, FASTA, PILEUP, FINDPATTERNS or the like (GCG, Madison, WI). Public sequence databases such as GenBank, EMBL, Swiss-Prot and PIR or private sequence databases such as PhytoSeq (Incyte Pharmaceuticals, Palo Alto, CA) may be searched. Alignment of sequences for comparison may be conducted by the local homology algorithm of Smith and Waterman (1981) *Adv. Appl. Math.* 2:482, by the homology alignment algorithm of Needleman and Wunsch (1970) *J. Mol. Biol.* 48:443, by the search for similarity method of Pearson and Lipman (1988) *Proc. Natl. Acad. Sci. U.S.A.* 85: 2444, by computerized implementations of these algorithms. After alignment, sequence comparisons between two (or more) polynucleotides or polypeptides are typically performed by comparing sequences of the two sequences over a comparison window to identify and compare local regions of sequence similarity. The comparison window may be a segment of at least about 20 contiguous positions, usually about 50 to about 200, more usually about 100 to about 150 contiguous positions. A description of the method is provided in Ausubel et al. (eds) (1999) *Current Protocols in Molecular Biology*, John Wiley & Sons.

Transcription factors that are homologs of the disclosed sequences will typically share at least 40% amino acid sequence identity. More closely related TFs may share at least 50%, 60%, 65%, 70%, 75% or 80% sequence identity with the disclosed sequences. Factors that are most closely related to the disclosed sequences share at least 85%, 90% or 95% sequence identity. At the nucleotide level, the sequences will typically share at least 40% nucleotide sequence identity, preferably at least 50%, 60%, 70% or 80% sequence identity, and more preferably 85%, 90%, 95% or 97% sequence identity. The degeneracy of the genetic code enables major variations in the nucleotide sequence of a polynucleotide while maintaining the amino acid sequence of the encoded protein.

One way to identify whether two nucleic acid molecules are closely related is that the two molecules hybridize to each other under stringent conditions. Generally, stringent conditions are selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Conditions for nucleic acid hybridization and calculation of stringencies can be found in Sambrook et al. (1989) *Molecular Cloning. A Laboratory Manual*, Ed. 2, Cold Spring Harbor Laboratory Press, New York and Tijssen (1993) *Laboratory Techniques in Biochemistry and*

Molecular Biology--Hybridization with Nucleic Acid Probes Part I, Elsevier, New York . Nucleic acid molecules that hybridize under stringent conditions will typically hybridize to a probe based on either the entire cDNA or selected portions of the cDNA under wash conditions of 0.2x SSC to 2.0 x SSC, 0.1% SDS at 50-65° C, for example 0.2 x SSC, 0.1% SDS at 65° C. For detecting less closely related homologs washes may be performed at 50° C.

For conventional hybridization the hybridization probe is conjugated with a detectable label such as a radioactive label, and the probe is preferably of at least 20 nucleotides in length. As is well known in the art, increasing the length of hybridization probes tends to give enhanced specificity. The labeled probe derived from the *Arabidopsis* nucleotide sequence may be hybridized to a plant cDNA or genomic library and the hybridization signal detected using means known in the art. The hybridizing colony or plaque (depending on the type of library used) is then purified and the cloned sequence contained in that colony or plaque isolated and characterized. Homologs may also be identified by PCR-based techniques, such as inverse PCR or RACE, using degenerate primers. See Ausubel et al. (eds) (1998) *Current Protocols in Molecular Biology*, John Wiley & Sons.

TF homologs may alternatively be obtained by immunoscreening an expression library. With the provision herein of the disclosed TF nucleic acid sequences, the polypeptide may be expressed and purified in a heterologous expression system (e.g., *E. coli*) and used to raise antibodies (monoclonal or polyclonal) specific for the TF. Antibodies may also be raised against synthetic peptides derived from TF amino acid sequences. Methods of raising antibodies are well known in the art and are described in Harlow and Lane (1988) *Antibodies: A Laboratory Manual*, Cold Spring Harbor Laboratory, New York. Such antibodies can then be used to screen an expression library produced from the plant from which it is desired to clone the TF homolog, using the methods described above. The selected cDNAs may be confirmed by sequencing and enzymatic activity.

3. Altered Expression of Transcription Factors

Any of the identified sequences may be incorporated into a cassette or vector for expression in plants. A number of expression vectors suitable for stable transformation of plant cells or for the establishment of transgenic plants have been described including those described in Weissbach and Weissbach, (1989) *Methods for Plant Molecular Biology*, Academic Press, and Gelvin et al., (1990) *Plant Molecular Biology Manual*, Kluwer Academic Publishers. Specific examples include those derived from a Ti plasmid of *Agrobacterium tumefaciens*, as well as those disclosed by Herrera-Estrella, L., et al., (1983) *Nature* 303: 209, Bevan, M., *Nucl. Acids Res.* (1984) 12: 8711-8721, Klee, H. J., (1985) *Bio/Technology* 3: 637-642, for dicotyledonous plants. Ti-derived plasmids can be transferred into both

monocotous and dicotyledonous species using *Agrobacterium*-mediated transformation (Ishida et al (1996) *Nat. Biotechnol.* 14:745-50; Barton et al. (1983) *Cell* 32:1033-1043).

Alternatively, non-Ti vectors can be used to transfer the DNA into plants and cells by using free DNA delivery techniques. Such methods may involve, for example, the use of liposomes, electroporation, microprojectile bombardment, silicon carbide whiskers, and viruses. By using these methods transgenic plants such as wheat, rice (Christou, P., (1991) *Bio/Technology* 9: 957-962) and corn (Gordon-Kamm, W., (1990) *Plant Cell* 2: 603-618) can be produced. An immature embryo can also be a good target tissue for monocots for direct DNA delivery techniques by using the particle gun (Weeks, T. et al., (1993) *Plant Physiol.* 102: 1077-1084; Vasil, V., (1993) *Bio/Technology* 10: 667-674; Wan, Y. and Lemeaux, P., (1994) *Plant Physiol.* 104: 37-48, and for *Agrobacterium*-mediated DNA transfer (Ishida et al., (1996) *Nature Biotech.* 14: 745-750).

Typically, plant transformation vectors include one or more cloned plant coding sequences (genomic or cDNA) under the transcriptional control of 5' and 3' regulatory sequences and a dominant selectable marker. Such plant transformation vectors typically also contain a promoter (e.g., a regulatory region controlling inducible or constitutive, environmentally-or developmentally-regulated, or cell- or tissue-specific expression), a transcription initiation start site, an RNA processing signal (such as intron splice sites), a transcription termination site, and/or a polyadenylation signal.

Examples of constitutive plant promoters which may be useful for expressing the TF sequence include: the cauliflower mosaic virus (CaMV) 35S promoter, which confers constitutive, high-level expression in most plant tissues (see, e.g., Odel et al., (1985) *Nature* 313:810); the nopaline synthase promoter (An et al., (1988) *Plant Physiol.* 88:547); and the octopine synthase promoter (Fromm et al., (1989) *Plant Cell* 1: 977).

A variety of plant gene promoters that regulate gene expression in response to environmental, hormonal, chemical, developmental signals, and in a tissue-active manner can be used for expression of the TFs in plants, as illustrated by seed-specific promoters (such as the napin, phaseolin or DC3 promoter described in US Pat. No. 5,773,697), root-specific promoters, such as those disclosed in US Patent Nos. 5,618,988, 5,837,848 and 5,905,186; fruit-specific promoters that are active during fruit ripening (such as the dru 1 promoter (US Pat. No. 5,783,393), or the 2A11 promoter (US Pat. No. 4,943,674) and the tomato polygalacturonase promoter (Bird et al. (1988) *Plant Mol. Biol.* 11:651), root-specific promoters, such as those disclosed in US Patent Nos. 5,618,988, 5,837,848 and 5,905,186, pollen-active promoters such as PTA29, PTA26 and PTA13 (US Pat. No. 5,792,929), promoters active in vascular tissue (Ringli and Keller (1998) *Plant Mol. Biol.* 37:977-988), flower-specific (Kaiser et al, (1995) *Plant Mol. Biol.* 28:231-243), pollen (Baerson et al. (1994) *Plant Mol. Biol.* 26:1947-1959), carpels (Ohl et al. (1990) *Plant Cell* 2:837-848), pollen and

ovules (Baerson et al. (1993) *Plant Mol. Biol.* 22:255-267) auxin-inducible promoters (such as that described in van der Kop et al (1999) *Plant Mol. Biol.* 39:979-990 or Baumann et al. (1999) *Plant Cell* 11:323-334), cytokinin-inducible promoter (Guevara-Garcia (1998) *Plant Mol. Biol.* 38:743-753), promoters responsive to gibberellin (Shi et al. (1998) *Plant Mol. Biol.* 38:1053-1060, Willmott et al. (1998) 38:817-825) and the like. Additional promoters are those that elicit expression in response to heat (Ainley, et al. (1993) *Plant Mol. Biol.* 22: 13-23), light (e.g., the pea rbcS-3A promoter, Kuhlemeier et al., (1989) *Plant Cell* 1:471, and the maize rbcS promoter, Schaffner and Sheen, (1991) *Plant Cell* 3: 997); wounding (e.g., *wun1*, Siebertz et al., (1989) *Plant Cell* 1: 961); pathogen resistance, and chemicals such as methyl jasmonate or salicylic acid (Gatz et al., (1997) *Plant Mol. Biol.* 48: 89-108). In addition, the timing of the expression can be controlled by using promoters such as those acting at late seed development (Odell et al. (1994) *Plant Physiol.* 106:447-458).

Plant expression vectors may also include RNA processing signals that may be positioned within, upstream or downstream of the coding sequence. In addition, the expression vectors may include additional regulatory sequences from the 3'-untranslated region of plant genes, e.g., a 3' terminator region to increase mRNA stability of the mRNA, such as the PI-II terminator region of potato or the octopine or nopaline synthase 3' terminator regions.

Finally, as noted above, plant expression vectors may also include dominant selectable marker genes to allow for the ready selection of transformants. Such genes include those encoding antibiotic resistance genes (e.g., resistance to hygromycin, kanamycin, bleomycin, G418, streptomycin or spectinomycin) and herbicide resistance genes (e.g., phosphinothricin acetyltransferase).

A reduction of TF expression in a transgenic plant to modify a plant trait may be obtained by introducing into plants antisense constructs based on the TF cDNA. For antisense suppression, the TF cDNA is arranged in reverse orientation relative to the promoter sequence in the expression vector. The introduced sequence need not be the full length TF cDNA or gene, and need not be identical to the TF cDNA or a gene found in the plant type to be transformed. Generally, however, where the introduced sequence is of shorter length, a higher degree of homology to the native TF sequence will be needed for effective antisense suppression. Preferably, the introduced antisense sequence in the vector will be at least 30 nucleotides in length, and improved antisense suppression will typically be observed as the length of the antisense sequence increases. Preferably, the length of the antisense sequence in the vector will be greater than 100 nucleotides. Transcription of an antisense construct as described results in the production of RNA molecules that are the reverse complement of mRNA molecules transcribed from the endogenous TF gene in the plant cell. Suppression of endogenous TF gene expression can also be achieved using a ribozyme. Ribozymes are

synthetic RNA molecules that possess highly specific endoribonuclease activity. The production and use of ribozymes are disclosed in U.S. Patent No. 4,987,071 to Cech and U.S. Patent No. 5,543,508 to Haselhoff. The inclusion of ribozyme sequences within antisense RNAs may be used to confer RNA cleaving activity on the antisense RNA, such that endogenous mRNA molecules that bind to the antisense RNA are cleaved, which in turn leads to an enhanced antisense inhibition of endogenous gene expression.

Vectors in which RNA encoded by the TF cDNA (or variants thereof) is over-expressed may also be used to obtain co-suppression of the endogenous TF gene in the manner described in U.S. Patent No. 5,231,020 to Jorgensen. Such co-suppression (also termed sense suppression) does not require that the entire TF cDNA be introduced into the plant cells, nor does it require that the introduced sequence be exactly identical to the endogenous TF gene. However, as with antisense suppression, the suppressive efficiency will be enhanced as (1) the introduced sequence is lengthened and (2) the sequence similarity between the introduced sequence and the endogenous TF gene is increased.

Vectors expressing an untranslatable form of the TF mRNA may also be used to suppress the expression of endogenous TF activity to modify a trait. Methods for producing such constructs are described in U.S. Patent No. 5,583,021 to Dougherty et al. Preferably, such constructs are made by introducing a premature stop codon into the TF gene. Alternatively, a plant trait may be modified by gene silencing using double-strand RNA (Sharp (1999) *Genes and Development* 13: 139-141).

Another method for abolishing the expression of a gene is by insertion mutagenesis using the T-DNA of *Agrobacterium tumefaciens*. After generating the insertion mutants, the mutants can be screened to identify those containing the insertion in a TF gene. Mutants containing a single mutation event at the desired gene may be crossed to generate homozygous plants for the mutation (Koncz et al. (1992) *Methods in Arabidopsis Research*. World Scientific).

A plant trait may also be modified by using the cre-lox system (for example, as described in US Pat. No. 5,658,772). A plant genome may be modified to include first and second lox sites that are then contacted with a Cre recombinase. If the lox sites are in the same orientation, the intervening DNA sequence between the two sites is excised. If the lox sites are in the opposite orientation, the intervening sequence is inverted.

The polynucleotides and polypeptides of this invention may also be expressed in a plant in the absence of an expression cassette by manipulating the activity or expression level of the endogenous gene by other means. For example, by ectopically expressing a gene by T-DNA activation tagging (Ichikawa et al., (1997) *Nature* 390 698-701, Kakimoto et al., (1996) *Science* 274: 982-985). This method entails transforming a plant with a gene tag containing multiple transcriptional enhancers and once the tag has inserted into the genome, expression of a flanking gene coding sequence becomes deregulated. In another example, the

transcriptional machinery in a plant may be modified so as to increase transcription levels of a polynucleotide of the invention (See PCT Publications WO9606166 and WO 9853057 which describe the modification of the DNA binding specificity of zinc finger proteins by changing particular amino acids in the DNA binding motif).

The transgenic plant may also comprise the machinery necessary for expressing or altering the activity of a polypeptide encoded by an endogenous gene, for example by altering the phosphorylation state of the polypeptide to maintain it in an activated state.

4. Transgenic Plants with Modified TF Expression

Once an expression cassette comprising a polynucleotide encoding a TF gene of this invention has been constructed, standard techniques may be used to introduce the polynucleotide into a plant in order to modify a trait of the plant. The plant may be any higher plant, including gymnosperms, monocotyledonous and dicotyledonous plants. Suitable protocols are available for *Leguminosae* (alfalfa, soybean, clover, etc.), *Umbelliferae* (carrot, celery, parsnip), *Cruciferae* (cabbage, radish, rapeseed, broccoli, etc.), *Curcubitaceae* (melons and cucumber), *Gramineae* (wheat, corn, rice, barley, millet, etc.), *Solanaceae* (potato, tomato, tobacco, peppers, etc.), and various other crops. See protocols described in Ammirato et al. (1984) *Handbook of Plant Cell Culture –Crop Species*. Macmillan Publ. Co. Shimamoto et al. (1989) *Nature* 338:274-276; Fromm et al. (1990) *Bio/Technology* 8:833-839; and Vasil et al. (1990) *Bio/Technology* 8:429-434.

Transformation and regeneration of both monocotyledonous and dicotyledonous plant cells is now routine, and the selection of the most appropriate transformation technique will be determined by the practitioner. The choice of method will vary with the type of plant to be transformed; those skilled in the art will recognize the suitability of particular methods for given plant types. Suitable methods may include, but are not limited to: electroporation of plant protoplasts; liposome-mediated transformation; polyethylene glycol (PEG) mediated transformation; transformation using viruses; micro-injection of plant cells; micro-projectile bombardment of plant cells; vacuum infiltration; and *Agrobacterium tumefaciens* mediated transformation. Transformation means introducing a nucleotide sequence in a plant in a manner to cause stable or transient expression of the sequence.

Successful examples of the modification of plant characteristics by transformation with cloned sequences which serve to illustrate the current knowledge in this field of technology, and which are herein incorporated by reference, include: U.S. Patent Nos. 5,571,706; 5,677,175; 5,510,471; 5,750,386; 5,597,945; 5,589,615; 5,750,871; 5,268,526; 5,780,708; 5,538,880; 5,773,269; 5,736,369 and 5,610,042.

Following transformation, plants are preferably selected using a dominant selectable marker incorporated into the transformation vector. Typically, such a marker will confer

antibiotic or herbicide resistance on the transformed plants, and selection of transformants can be accomplished by exposing the plants to appropriate concentrations of the antibiotic or herbicide.

After transformed plants are selected and grown to maturity, those plants showing a modified trait are identified. The modified trait may be any of those traits described above. Additionally, to confirm that the modified trait is due to changes in expression levels or activity of the polypeptide or polynucleotide of the invention may be determined by analyzing mRNA expression using Northern blots, RT-PCR or microarrays, or protein expression using immunoblots or Western blots or gel shift assays.

The plants may have commercial utility for increasing tolerance or resistance to pathogens and pests. These transgenic plants may be more resistant to biotrophic or necrotrophic pathogens or belonging to the following groups such as a fungus, bacterium, mollicute, virus, nematode, a parasitic higher plant or the like and associated diseases. In particular, pathogens such as *Fusarium oxysporum*, *Erysiphe orontii* and other powdery mildews, *Sclerotinia spp.*, soil-borne oomycetes, foliar oomycetes, *Botrytis spp.*, *Rhizoctonia spp.*, *Verticillium dahliae/albo-atrum*, *Alternaria spp.*, rusts, *Mycosphaerella spp.*, *Fusarium solani*, or the like. The diseases include fungal diseases such as rusts, smuts, wilts, yellows, root rot, leaf drop, ergot, leaf blight of potato, brown spot of rice, leaf blight, late blight, powdery mildew, downy mildew, and the like; viral diseases such as sugarcane mosaic, cassava mosaic, sugar beet yellows, plum pox, barley yellow dwarf, tomato yellow leaf curl, tomato spotted wilt virus, and the like; bacterial diseases such as citrus canker, bacterial leaf blight, bacterial wilt, soft rot of vegetables, and the like; nematode diseases caused by parasitic nematodes such as root-knot nematodes, cyst nematodes or the like.

5. Other Utility of the Polypeptide and Polynucleotide Sequences

A transcription factor provided by the present invention may also be used to identify exogenous or endogenous molecules that may affect expression of the transcription factors and may affect any of the traits described herein. These molecules may include organic or inorganic compounds.

For example, the method may entail first placing the molecule in contact with a plant or plant cell. The molecule may be introduced by topical administration, such as spraying or soaking of a plant, and then the molecule's effect on the expression or activity of the TF polypeptide or the expression of the polynucleotide monitored. Changes in the expression of the TF polypeptide may be monitored by use of polyclonal or monoclonal antibodies, gel electrophoresis or the like. Changes in the expression of the corresponding polynucleotide sequence may be detected by use of microarrays, Northern blots or any other technique for monitoring changes in mRNA expression. These techniques are exemplified in Ausubel et al.

(eds) *Current Protocols in Molecular Biology*, John Wiley & Sons (1998). Such changes in the expression levels may be correlated with modified plant traits and thus identified molecules may be useful for soaking or spraying on fruit, vegetable and grain crops to modify traits in plants.

5 The transcription factors may also be employed to identify promoter sequences with which they may interact. After identifying a promoter sequence, interactions between the transcription factor and the promoter sequence may be modified by changing specific nucleotides in the promoter sequence or specific amino acids in the transcription factor that interact with the promoter sequence to alter a plant trait. Typically, transcription factor DNA
10 binding sites are identified by gel shift assays. After identifying the promoter regions, the promoter region sequences may be employed in double-stranded DNA arrays to identify molecules that affect the interactions of the TFs with their promoters (Bulyk et al. (1999) *Nature Biotechnology* 17:573-577).

15 The identified transcription factors are also useful to identify proteins that modify the activity of the transcription factor. Such modification may occur by covalent modification, such as by phosphorylation, or by protein-protein (homo or-heteropolymer) interactions. Any method suitable for detecting protein-protein interactions may be employed. Among the methods that may be employed are co-immunoprecipitation, cross-linking and co-purification through gradients or chromatographic columns, and the two-hybrid yeast system.

20 The two-hybrid system detects protein interactions in vivo and is described in Chien, et al., (1991), *Proc. Natl. Acad. Sci. USA*, 88, 9578-9582 and is commercially available from Clontech (Palo Alto, Calif.). In such a system, plasmids are constructed that encode two hybrid proteins: one consists of the DNA-binding domain of a transcription activator protein fused to the TF polypeptide and the other consists of the transcription activator protein's
25 activation domain fused to an unknown protein that is encoded by a cDNA that has been recombined into the plasmid as part of a cDNA library. The DNA-binding domain fusion plasmid and the cDNA library are transformed into a strain of the yeast *Saccharomyces cerevisiae* that contains a reporter gene (e.g., lacZ) whose regulatory region contains the transcription activator's binding site. Either hybrid protein alone cannot activate transcription of
30 the reporter gene. Interaction of the two hybrid proteins reconstitutes the functional activator protein and results in expression of the reporter gene, which is detected by an assay for the reporter gene product. Then, the library plasmids responsible for reporter gene expression are isolated and sequenced to identify the proteins encoded by the library plasmids. After identifying proteins that interact with the transcription factors, assays for compounds that
35 interfere with the TF protein-protein interactions may be preformed.

The following examples are intended to illustrate but not limit the present invention.

Example I. Full Length Gene Identification and Cloning

Putative transcription factor sequences (genomic or ESTs) related to known transcription factors were identified in the *Arabidopsis thaliana* GenBank database using the tblastn sequence analysis program using default parameters and a P-value cutoff threshold of -4 or -5 or lower, depending on the length of the query sequence. Putative transcription factor sequence hits were then screened to identify those containing particular sequence strings. If the sequence hits contained such sequence strings, the sequences were confirmed as transcription factors.

Alternatively, *Arabidopsis thaliana* cDNA libraries derived from different tissues or treatments, or genomic libraries were screened to identify novel members of a transcription family using a low stringency hybridization approach. Probes were synthesized using gene specific primers in a standard PCR reaction (annealing temperature 60° C) and labeled with ³²P dCTP using the High Prime DNA Labeling Kit (Boehringer Mannheim). Purified radiolabelled probes were added to filters immersed in Church hybridization medium (0.5 M NaPO₄ pH 7.0, 7% SDS, 1 % w/v bovine serum albumin) and hybridized overnight at 60 °C with shaking. Filters were washed two times for 45 to 60 minutes with 1xSSC, 1% SDS at 60° C.

To identify additional sequence 5' or 3' of a partial cDNA sequence in a cDNA library, 5' and 3' rapid amplification of cDNA ends (RACE) was performed using the Marathon™ cDNA amplification kit (Clontech, Palo Alto, CA). Generally, the method entailed first isolating poly(A) mRNA, performing first and second strand cDNA synthesis to generate double stranded cDNA, blunting cDNA ends, followed by ligation of the Marathon™ Adaptor to the cDNA to form a library of adaptor-ligated ds cDNA. Gene-specific primers were designed to be used along with adaptor specific primers for both 5' and 3' RACE reactions. Nested primers, rather than single primers, were used to increase PCR specificity. Using 5' and 3' RACE reactions, 5' and 3' RACE fragments were obtained, sequenced and cloned. The process may be repeated until 5' and 3' ends of the full-length gene were identified. Then the full-length cDNA was generated by PCR using primers specific to 5' and 3' ends of the gene by end-to-end PCR.

Example II Pathogen Resistance Genes

RT-PCR and microarray experiments were performed to identify those genes induced after exposure to biotrophic fungal pathogens, such as *Erysiphe orontii*, necrotrophic fungal pathogens, such as *Fusarium oxysporum*, and disease associated growth-regulators such as salicylic acid, methyl jasmonate and ethylene (ACC). The gene expression patterns from soil grown as well as tissue culture grown plant tissue were investigated.

Fusarium oxysporum isolates cause vascular wilts and damping off of various annual vegetables, perennials and weeds (Mauch-Mani and Slusarenko (1994) Molecular Plant-Microbe Interactions 7: 378-383). For *Fusarium oxysporum* experiments, plants grown on petri dishes were sprayed with a fresh spore suspension of *F. oxysporum*. The spore suspension was prepared as follows: A plug of fungal hyphae from a plate culture was placed on a fresh potato dextrose agar plate and allowed to spread for one week. 5 ml sterile water was then added to the plate, swirled, and pipetted into 50 ml Armstrong Fusarium medium. Spores were grown overnight in Fusarium medium and then sprayed onto plants using a Preval paint sprayer. Plant tissue was harvested and frozen in liquid nitrogen 48 hours post infection

Erysiphe orontii is a causal agent of powdery mildew. For *Erysiphe orontii* experiments, plants were grown approximately 4 weeks in a greenhouse under 12 hour light (20 C, ~30% relative humidity (rh)). Individual leaves were infected with *E. orontii* spores from infected plants using a camel's hair brush, and the plants were transferred to a Percival growth chamber (20 C, 80% rh.). Plant tissue was harvested and frozen in liquid nitrogen 7 days post infection.

For salicylic acid experiments, 15 day old seedlings grown on petri dishes were transferred to plates containing 0.5 mM salicylic acid (SA). After 72 hours, leaves were harvested and frozen in liquid nitrogen.

Reverse transcriptase PCR was done using gene specific primers within the coding region for each sequence identified. The primers were designed near the 3' region of each coding sequence initially identified.

Total RNA from these tissues were isolated using the CTAB extraction protocol. Once extracted total RNA was normalized in concentration across all the tissue types to ensure that the PCR reaction for each tissue received the same amount of cDNA template using the 28S band as reference. Poly A+ was purified using a modified protocol from the Qiagen Oligotex kit batch protocol. cDNA was synthesized using standard protocols. After the first strand cDNA synthesis, primers for Actin 2 were used to normalize the concentration of cDNA across the tissue types. Actin 2 is found to be constitutively expressed in fairly equal levels across the tissue types we are investigating.

For RT PCR, cDNA template was mixed with corresponding primers and Taq polymerase. Each reaction consisted of 0.2 ul cDNA template, 2ul 10X Tricine buffer, 2 ul 10X Tricine buffer and 16.8 ul water, 0.05ul Primer 1, 0.05 ul, Primer 2, 0.3 ul Taq polymerase and 8.6 ul water.

The 96 well plate was covered with microfilm and set in the Thermocycler to start the following reaction cycle. Step1 93° C for 3 mins, Step 2 93° C for 30 sec, Step 3 65° C for 1 min, Step 4 72° C for 2 mins,. Steps 2, 3 and 4 were repeated for 28 cycles, Step 5 72° C

for 5 mins and Step 6 4° C. The PCR plate was placed back in the thermocycler to amplify more products at 8 more cycles to identify genes that have very low expression. The reaction cycle was as follows: Step 2 93° C for 30 sec, Step 3 65° C for 1 min, and Step 4 72° C for 2 ins, repeated for 8 cycles, and Step 4 4° C.

8ul of PCR product and 1.5 ul of loading dye were loaded on a 1.2% agarose gel for analysis after 28 cycles and 36 cycles. Expression levels of specific transcripts were considered low if they were only detectable after 36 cycles of PCR. Expression levels were considered medium or high depending on the levels of transcript compared with observed transcript levels for actin2.

In some instances, expression patterns of the transcription factors was monitored by microarray experiments. cDNAs were generated by PCR and resuspended at a final concentration of ~ 100 ng/ul in 3X SSC or 150mM Na-phosphate (Eisen and Brown (1999) *Meth. in Enzymol.* 303:179-205). The cDNAs were spotted on microscope glass slides coated with polylysine. The prepared cDNAs were aliquoted into 384 well plates and spotted on the slides using an x-y-z gantry (OmniGrid) purchased from GeneMachines (Menlo Park, CA) outfitted with quill type pins purchased from Telechem International (Sunnyvale, CA). After spotting, the arrays were cured for a minimum of one week at room temperature, rehydrated and blocked following the protocol recommended by Eisen and Brown (1999).

Sample total RNA (10 ug) samples were labeled using fluorescent Cy3 and Cy5 dyes. Labeled samples were resuspended in 4X SSC/0.03% SDS/4 ug salmon sperm DNA/2 ug tRNA/ 50mM Na-pyrophosphate, heated for 95°C for 2.5 minutes, spun down and placed on the array. The array was then covered with a glass coverslip and placed in a sealed chamber. The chamber was then kept in a water bath at 62°C overnight. The arrays were washed as described in Eisen and Brown (1999) and scanned on a General Scanning 3000 laser scanner. The resulting files are subsequently quantified using Imagene a software purchased from BioDiscovery (Los Angeles, CA).

The transcript levels were observed to be upregulated between 1.5 and 100 fold when compared with control plants not exposed to the pathogens.

Example III. Construction of Expression Vectors

The sequence was amplified from a genomic or cDNA library using primers specific to sequences upstream and downstream of the coding region. The expression vector was pMEN20, which is derived from pMON316 (Sanders et al, (1987) *Nucleic Acids Research* 15:1543-58). To clone the sequence into the vector, both pMEN20 and the amplified DNA fragment were digested separately with Sall and NotI restriction enzymes at 37° C for 2 hours. The digestion products were subject to electrophoresis in a 0.8% agarose gel and visualized

by ethidium bromide staining. The DNA fragments containing the sequence and the linearized plasmid were excised and purified by using a Qiaquick gel extraction kit (Qiagen, CA). The fragments of interest were ligated at a ratio of 3:1 (vector to insert). Ligation reactions using T4 DNA ligase (New England Biolabs, MA) were carried out at 16° C for 16 hours. The ligated DNAs were transformed into competent cells of the *E. coli* strain DH5alpha by using the heat shock method. The transformations were plated on LB plates containing 50 mg/l spectinomycin (Sigma).

Individual colonies were grown overnight in five milliliters of LB broth containing 50 mg/l spectinomycin at 37° C. Plasmid DNA was purified by using Qiaquick Mini Prep kits (Qiagen, CA).

Example IV. Transformation of *Agrobacterium* with the Expression Vector

After the plasmid vector containing the gene was constructed, the vector was used to transform *Agrobacterium tumefaciens* cells expressing the gene products. The stock of *Agrobacterium tumefaciens* cells for transformation were made as described by Nagel et al. *FEMS Microbiol Letts* 67: 325-328 (1990). *Agrobacterium* strain GV3101 was grown in 250 ml LB medium (Sigma) overnight at 28°C with shaking until an absorbance (A_{600}) of 0.5 – 1.0 was reached. Cells were harvested by centrifugation at 4,000 x g for 15 min at 4° C. Cells were then resuspended in 250 µl chilled buffer (1 mM HEPES, pH adjusted to 7.0 with KOH). Cells were centrifuged again as described above and resuspended in 125 µl chilled buffer. Cells were then centrifuged and resuspended two more times in the same HEPES buffer as described above at a volume of 100 µl and 750 µl, respectively. Resuspended cells were then distributed into 40 µl aliquots, quickly frozen in liquid nitrogen, and stored at -80° C.

Agrobacterium cells were transformed with plasmids prepared as described above following the protocol described by Nagel et al. *FEMS Microbiol Letts* 67: 325-328 (1990). For each DNA construct to be transformed, 50 – 100 ng DNA (generally resuspended in 10 mM Tris-HCl, 1 mM EDTA, pH 8.0) was mixed with 40 µl of *Agrobacterium* cells. The DNA/cell mixture was then transferred to a chilled cuvette with a 2mm electrode gap and subject to a 2.5 kV charge dissipated at 25 µF and 200 µF using a Gene Pulser II apparatus (Bio-Rad). After electroporation, cells were immediately resuspended in 1.0 ml LB and allowed to recover without antibiotic selection for 2 – 4 hours at 28° C in a shaking incubator. After recovery, cells were plated onto selective medium of LB broth containing 100 µg/ml spectinomycin (Sigma) and incubated for 24-48 hours at 28° C. Single colonies were then picked and inoculated in fresh medium. The presence of the plasmid construct was verified by PCR amplification and sequence analysis.

Example V. Transformation of *Arabidopsis* Plants with *Agrobacterium tumefaciens* with Expression Vector

After transformation of *Agrobacterium tumefaciens* with plasmid vectors containing the gene, single *Agrobacterium* colonies were identified, propagated, and used to transform *Arabidopsis* plants. Briefly, 500 ml cultures of LB medium containing 50 mg/l spectinomycin were inoculated with the colonies and grown at 28° C with shaking for 2 days until an absorbance (A_{600}) of > 2.0 is reached. Cells were then harvested by centrifugation at 4,000 x g for 10 min, and resuspended in infiltration medium (1/2 X Murashige and Skoog salts (Sigma), 1 X Gamborg's B-5 vitamins (Sigma), 5.0% (w/v) sucrose (Sigma), 0.044 μ M benzylamino purine (Sigma), 200 μ L/L Silwet L-77 (Lehle Seeds) until an absorbance (A_{600}) of 0.8 was reached.

Prior to transformation, *Arabidopsis thaliana* seeds (ecotype Columbia) were sown at a density of ~10 plants per 4" pot onto Pro-Mix BX potting medium (Hummert International) covered with fiberglass mesh (18 mm X 16 mm). Plants were grown under continuous illumination (50-75 μ E/m²/sec) at 22-23° C with 65-70% relative humidity. After about 4 weeks, primary inflorescence stems (bolts) are cut off to encourage growth of multiple secondary bolts. After flowering of the mature secondary bolts, plants were prepared for transformation by removal of all siliques and opened flowers.

The pots were then immersed upside down in the mixture of *Agrobacterium* infiltration medium as described above for 30 sec, and placed on their sides to allow draining into a 1' x 2' flat surface covered with plastic wrap. After 24 h, the plastic wrap was removed and pots are turned upright. The immersion procedure was repeated one week later, for a total of two immersions per pot. Seeds were then collected from each transformation pot and analyzed following the protocol described below.

Example VI. Identification of *Arabidopsis* Primary Transformants

Seeds collected from the transformation pots were sterilized essentially as follows. Seeds were dispersed into in a solution containing 0.1% (v/v) Triton X-100 (Sigma) and sterile H₂O and washed by shaking the suspension for 20 min. The wash solution was then drained and replaced with fresh wash solution to wash the seeds for 20 min with shaking. After removal of the second wash solution, a solution containing 0.1% (v/v) Triton X-100 and 70% ethanol (Equistar) was added to the seeds and the suspension was shaken for 5 min. After removal of the ethanol/detergent solution, a solution containing 0.1% (v/v) Triton X-100 and 30% (v/v) bleach (Clorox) was added to the seeds, and the suspension was shaken for 10 min. After removal of the bleach/detergent solution, seeds were then washed five times in sterile distilled H₂O. The seeds were stored in the last wash water at 4° C for 2 days in the dark before being plated onto antibiotic selection medium (1 X Murashige and Skoog salts (pH

adjusted to 5.7 with 1M KOH), 1 X Gamborg's B-5 vitamins, 0.9% phytagar (Life Technologies), and 50 mg/l kanamycin). Seeds were germinated under continuous illumination (50-75 $\mu\text{E}/\text{m}^2/\text{sec}$) at 22-23° C. After 7-10 days of growth under these conditions, kanamycin resistant primary transformants (T_1 generation) were visible and obtained. These seedlings were transferred first to fresh selection plates where the seedlings continued to grow for 3-5 more days, and then to soil (Pro-Mix BX potting medium).

Primary transformants are self-crossed and progeny seeds (T_2) collected.

Example VII. Analysis of Arabidopsis T_2 progeny plants for Pathogen Resistance or Pathogen Tolerance

T_2 or knockout mutant seed were surface sterilized and sown on MS media containing sucrose. Ten days post-planting, seedlings were transferred to MS media without sucrose. At two weeks of age *Arabidopsis* seedlings were inoculated with *Fusarium* by spraying with a spore suspension (2×10^6 conidia per milliliter) and incubated under high humidity. Plants were then scored macroscopically for disease symptoms or microscopically for fungal growth or using microarrays for the induction of resistance associated genes (such as the defensin genes) to detect resistance or tolerance of the plant tissue. A wild type plant shows the first signs of damage (gradual yellowing of leaves, damping off of seedlings or growth of fungal mycelium) after two to four days after inoculation. Transgenic plants that are pathogen resistant or tolerant showed a delay in disease or symptom development compared to wild-type control plants.

Alternatively, *Erysiphe* inoculations were done by tapping conidia from 1 to 2 heavily infected leaves onto the mesh cover of a settling tower, brushing the mesh with a camel's hair paint brush to break up the conidial chains, and letting the conidia settle for 10 minutes. Plants were 4 to 4.5 weeks old at the time of inoculation. Spores were obtained from 10 to 14 day old *Erysiphe* cultures. Typically, within the first twenty-four hours, the spores differentiated into several fungal structures including the haustorium that invaginates a host's epidermal plasma membrane. Formation of aerial mycelium and sporulation represent late differentiation events between 4 and 7 days post inoculation (Freilaldenhoven et al. (1994) *Plant Cell* 6: 983-994). Plant resistance was scored based on the relative number and size of mycelial patches bearing conidia compared to wild-type control plants. Events associated with disease resistance to the pathogens and pests include: the induction of pathogen resistance related genes (R genes), the activation of cell death in the attacked epidermal cells (hypersensitive response), the induction of anti-microbial compounds, such as phytoalexins, and the lignification that occurs at attempted penetration sites. Assays are performed to observe these events. Transgenic plants identified that induce R genes, activate cell death, induce anti-microbial compounds or increase lignification sooner or to a greater extent than

wild-type plants when exposed to pathogen are potentially more resistant to infection by *Erysiphe* as well as a number of other pathogens and pests.

We have observed that when the expression levels of the genes are altered, that the disease phenotype can be varied. For example, G19 was significantly induced upon infection by the fungal pathogen *Erysiphe orontii* as well as the disease associated growth regulator, ethylene. Our data show that G19 overexpressing plants were more tolerant to infection with a moderate dose of *Erysiphe orontii* and in a nematode screen. The transgenic plants overexpressing G19 under the control of the 35S promoter were morphologically similar to control plants.

Additionally, G511 was another example of a gene that when overexpressed showed an increased tolerance to the fungal pathogen *Erysiphe orontii*. In both cases increased tolerance includes a significant reduction in pathogen growth and symptom development compared to wild type plants that were treated with pathogen in an identical manner.

Example VIII. Transformation of Cereal Plants with the Expression Vector

A cereal plant, such as corn, wheat, rice, sorghum or barley, can also be transformed with the plasmid vectors containing the sequence and constitutive or inducible promoters to modify a trait. In these cases, a cloning vector, pMEN020, is modified to replace the NptII coding region with the BAR gene of *Streptomyces hygroscopicus* that confers resistance to phosphinothricin. The KpnI and BglII sites of the Bar gene are removed by site-directed mutagenesis with silent codon changes.

Plasmids according to the present invention may be transformed into corn embryogenic cells derived from immature scutellar tissue by using microprojectile bombardment, with the A188XB73 genotype as the preferred genotype (Fromm et al., *Bio/Technology* 8: 833-839 (1990); Gordon-Kamm et al., *Plant Cell* 2: 603-618 (1990)). After microprojectile bombardment the tissues are selected on phosphinothricin to identify the transgenic embryogenic cells (Gordon-Kamm et al., *Plant Cell* 2: 603-618 (1990)). Transgenic plants are regenerated by standard corn regeneration techniques (Fromm, et al., *Bio/Technology* 8: 833-839 (1990); Gordon-Kamm et al., *Plant Cell* 2: 603-618 (1990)).

Example IX. Identification of Homologous Sequences

Homologs from the same plant, different plant species or other organisms were identified using database sequence search tools, such as the Basic Local Alignment Search Tool (BLAST) (Altschul et al. (1990) *J. Mol. Biol.* 215:403-410; and Altschul et al. (1997) *Nucl. Acid Res.* 25: 3389-3402). The tblastn or blastn sequence analysis programs were employed using the BLOSUM-62 scoring matrix (Henikoff, S. and Henikoff, J. G. (1992) *Proc. Natl. Acad. Sci. USA* 89: 10915-10919). The output of a BLAST report provides a score that takes

into account the alignment of similar or identical residues and any gaps needed in order to align the sequences. The scoring matrix assigns a score for aligning any possible pair of sequences. The P values reflect how many times one expects to see a score occur by chance. Higher scores are preferred and a low threshold P value threshold is preferred.

These are the sequence identity criteria. The tblastn sequence analysis program was used to query a polypeptide sequence against six-way translations of sequences in a nucleotide database. Hits with a P value less than -25, preferably less than -70, and more preferably less than -100, were identified as homologous sequences (exemplary selected sequence criteria). The blastn sequence analysis program was used to query a nucleotide sequence against a nucleotide sequence database. In this case too, higher scores were preferred and a preferred threshold P value was less than -13, preferably less than -50, and more preferably less than -100.

Alternatively, a fragment of a sequence from Figure 1 is ³²P-radiolabeled by random priming (Sambrook et al., (1989) *Molecular Cloning. A Laboratory Manual*, 2nd Ed., Cold Spring Harbor Laboratory Press, New York) and used to screen a plant genomic library (the exemplary test polynucleotides). As an example, total plant DNA from *Arabidopsis thaliana*, *Nicotiana tabacum*, *Lycopersicon pimpinellifolium*, *Prunus avium*, *Prunus cerasus*, *Cucumis sativus*, or *Oryza sativa* are isolated according to Stockinger al (Stockinger, E. J., et al., (1996), *J. Heredity*, 87:214-218). Approximately 2 to 10 µg of each DNA sample are restriction digested, transferred to nylon membrane (Micron Separations, Westboro, MA) and hybridized. Hybridization conditions are: 42° C in 50% formamide, 5X SSC, 20 mM phosphate buffer 1X Denhardt's, 10% dextran sulfate, and 100µg/ml herring sperm DNA. Four low stringency washes at RT in 2X SSC, 0.05% sodium sarcosyl and 0.02% sodium pyrophosphate are performed prior to high stringency washes at 55° C in 0.2X SSC, 0.05% sodium sarcosyl and 0.01% sodium pyrophosphate. High stringency washes are performed until no counts are detected in the washout according to Walling et al. (Walling, L. L., et al., (1988) *Nucl. Acids Res.* 16:10477-10492).

All references (publications and patents) are incorporated herein by reference in their entirety for all purposes.

Although the invention has been described with reference to the embodiments and examples above, it should be understood that various modifications can be made without departing from the spirit of the invention. Accordingly, the invention is limited only by the following claims.